

WANLI: Worker and AI Collaboration for Natural Language Inference Dataset Creation

Year: 2022 | Citations: 247 | Authors: Alisa Liu, Swabha Swayamdipta, Noah A. Smith, Yejin Choi

Abstract

A recurring challenge of crowdsourcing NLP datasets at scale is that human writers often rely on repetitive patterns when crafting examples, leading to a lack of linguistic diversity. We introduce a novel approach for dataset creation based on worker and AI collaboration, which brings together the generative strength of language models and the evaluative strength of humans. Starting with an existing dataset, MultiNLI for natural language inference (NLI), our approach uses dataset cartography to automatically identify examples that demonstrate challenging reasoning patterns, and instructs GPT-3 to compose new examples with similar patterns. Machine generated examples are then automatically filtered, and finally revised and labeled by human crowdworkers. The resulting dataset, WANLI, consists of 107,885 NLI examples and presents unique empirical strengths over existing NLI datasets. Remarkably, training a model on WANLI improves performance on eight out-of-domain test sets we consider, including by 11% on HANS and 9% on Adversarial NLI, compared to training on the 4x larger MultiNLI. Moreover, it continues to be more effective than MultiNLI augmented with other NLI datasets. Our results demonstrate the promise of leveraging natural language generation techniques and re-imagining the role of humans in the dataset creation process.